

2590
0574

7



OIPE

ENTERED

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/027,000

TIME: 09:51:58

Input Set : A:\GC696-SEQLIST.txt

Output Set: N:\CRF3\05102002\J027000.raw

```

4 <110> APPLICANT: Dunn-Coleman, Nigel
5   Goedegebuur, Frits
6   Ward, Michael
7   Yao, Jian
9 <120> TITLE OF INVENTION: BGL4 Beta-Glucosidase and Nucleic Acids
10  Encoding the Same
12 <130> FILE REFERENCE: GC696
14 <140> CURRENT APPLICATION NUMBER: US 10/027,000
15 <141> CURRENT FILING DATE: 2001-12-18
17 <160> NUMBER OF SEQ ID NOS: 3
19 <170> SOFTWARE: FastSEQ for Windows Version 4.0
21 <210> SEQ ID NO: 1
22 <211> LENGTH: 2976
23 <212> TYPE: DNA
24 <213> ORGANISM: Trichoderma reesei
26 <400> SEQUENCE: 1
27 ttatagtcgc tttgttaa at tggcctcgag gtcgaccac gcgtccggct gcttgtcccg      60
28 ttctgtgcc tgatgtctat ctgccgttg cctcctcatc ctcatctccc tgttgtctgt      120
29 ctctcttag tgcttcagtg acgctagggt cggtcacttt gtccccctt cgttgctctg      180
30 gtgtgtccaa ggtctaccct gcagtggttt tgaacccttg atatcctgct tgagcatccg      240
31 cgctgccata tagagcagca tatttttcta tctccaaaga tcccctcacc gagagttcta      300
32 ttaccccgac ccttgccctg tcatccagtc cttccatcat ggctgatatt gatgttgagg      360
33 ccatcttgaa gaagctcacc ctggccgaga aggtcgatct gctggctggt atcgacttct      420
34 ggcacacaaa ggcctctccc aagcatggag tcccctctct ccgctttaca gatggcccca      480
35 acggcgtaag agggaccaag ttcttcaatg gcgtccctgc ggctgcttc ccttgccgga      540
36 cgctcgctcg ttccacattc aaccaaatc tgctcgaaga ggcaggtaag atgatgggca      600
37 aagaggccat cgctaagagt gcgcatgtga tccctggccc gactatcaac atgcaacgct      660
38 cccctctcgg tggacgtggc ttcgagtcga ttggtgagga tccgttcctg gcgggcttg      720
39 gagctgcggc tctcatccgc ggcattcaga gcactggagt gcaggctacg atcaagcact      780
40 ttttgtgcaa tgatcaggag gacaggcgca tgatggtgca gagcatcgtc acggagcggg      840
41 ctctccgtga aatctacgca ctcccgttcc agattgctgt gcgagactcc cagccgggtg      900
42 cgttcatgac ggcgtacaat ggcattcaat gcgtgtcgtg cagcgagaac cctaaatatc      960
43 ttgatgggat gcttcgaaag gaatgggggt gggatggcct aatcatgagc gactggtacg     1020
44 gcacatacag taccacagaa gccgttgttg caggcctcga cctcgagatg cccggacctc     1080
45 cacgcttccg aggagaaaca ctcaagttca acgtctccaa cggaaagccc tttatccacg     1140
46 tcattgacca gagggctagg gaagttcttc agttcgtcaa gaagtgtgct gcctccggag     1200
47 tgacggagaa cggccccgag acgactgtca acaacacccc cgaaacggca gctctcctcc     1260
48 ggaaggttgg caacgagggc atcgtgctgc tgaagaacga gaacaacgtt ctgcccttga     1320
49 gcaagaagaa gaagacgctg attgtcggcc ccaacgccaa gcaggccaca taccacggcg     1380
50 gaggctctgc cgcactcagg gcctactacg cagtcactcc ctttgacggc ctacgaagc     1440
51 agctcgagag cccgcatcag tacaccgtcg gcgcctacac caccgttcct cccattctag     1500
52 gcgagcagtg cctcacgccc gacggcgctc cgggcatgcg ctggagggtc ttcaacgagc     1560
53 cccctggtac ccctaaccgc cagcacattg acgagctctt cttaccaag acggacatgc     1620

```

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/027,000

TIME: 09:51:58

Input Set : A:\GC696-SEQLIST.txt

Output Set : N:\CRF3\05102002\J027000.raw

```

54 acctggtgga ctactaccac cccaaggcgg cagacacgtg gtacgccgac atggagggca 1680
55 cgtacaccgc cgacgaggac tgcacctacg agctcgccct cgtcgtctgc ggcacggcaa 1740
56 aggcgtacgt agacgaccag ctctcgtcgc acaacgccac caagcaggtc cccggcgatg 1800
57 ccttcttcgg ctccgccacc cgcgaggaga cgggccgcat caatctcgtc aagggaaca 1860
58 cgtacaagtt caagatcgag ttcggtcccg caccaccta caccctcaag ggcgacacca 1920
59 tcgtccccgg ccacggctcc ctccgcgtcg gcggtgcaa ggtcattgac gaccaggccg 1980
60 aaatcgaaaa gtccgtcgcc ctcgccaagg agcacgacca ggtcatcatc tgcgcgggcc 2040
61 ttaacgccga ctgggagacc gagggcgccg accgcgcgag catgaagtc cccggcgtgc 2100
62 tggaccagct cattgccgac gtggccgccc cgaacccaaa caccgtcgtc gtcattgcaga 2160
63 cgggcacccc cgaggagatg ccctggctcg acgccacgcc cgccgtcatc caggcctggt 2220
64 acggcggcaa cgagacgggc aactccattg ccgacgtcgt ctttggcgac tacaaccct 2280
65 cgggcaagct gtccctcagc ttccccaagc gcctgcagga caaccccgcg tttctcaact 2340
66 tccgcaccga ggccggggcg acgctgtacg gcgaggacgt ctacgtcggg tacagggtact 2400
67 acgagtttgc cgacaaggac gtcaatttcc cctttggcca cggcctgtcc tacaccactt 2460
68 ttgccttttc caatctctcc gtgtctcaca aggacggcaa gctgagcgtg tccctctccg 2520
69 tgaagaacac cggtccgtg cccggcgcac aggtggccca gctctacgtc aagccctcc 2580
70 aagcggccaa gattaaccgc cccgtcaagg agtcaaggg cttcgcaaag gtcgaactgc 2640
71 agcccgcgga gacgaaggcg gtgacaatcg aggagcagga gaagtacgtc gctgcgtatt 2700
72 ttgatgagga gcgggatcag tgggtgtcgc aaaagggtga ctatgaggtt atcgtgagcg 2760
73 acagcagcgc agcgaaggat ggggttgccg tcaggggtaa gtttacggtg ggagagacgt 2820
74 attggtggtc tggcgtgtaa agtcgtgcat catctttggc agattgaatc cagtcacttt 2880
75 gcatatagcg ccgatgaaat gagaaccaac gatcattgtg taacaaaaaa aaaaaaaa 2940
76 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa 2976

```

78 <210> SEQ ID NO: 2

79 <211> LENGTH: 833

80 <212> TYPE: PRT

81 <213> ORGANISM: Trichoderma reesei

83 <400> SEQUENCE: 2

```

84 Met Ala Asp Ile Asp Val Glu Ala Ile Leu Lys Lys Leu Thr Leu Ala
85 1 5 10 15
86 Glu Lys Val Asp Leu Leu Ala Gly Ile Asp Phe Trp His Thr Lys Ala
87 20 25 30
88 Leu Pro Lys His Gly Val Pro Ser Leu Arg Phe Thr Asp Gly Pro Asn
89 35 40 45
90 Gly Val Arg Gly Thr Lys Phe Phe Asn Gly Val Pro Ala Ala Cys Phe
91 50 55 60
92 Pro Cys Gly Thr Ser Leu Gly Ser Thr Phe Asn Gln Thr Leu Leu Glu
93 65 70 75 80
94 Glu Ala Gly Lys Met Met Gly Lys Glu Ala Ile Ala Lys Ser Ala His
95 85 90 95
96 Val Ile Leu Gly Pro Thr Ile Asn Met Gln Arg Ser Pro Leu Gly Gly
97 100 105 110
98 Arg Gly Phe Glu Ser Ile Gly Glu Asp Pro Phe Leu Ala Gly Leu Gly
99 115 120 125
100 Ala Ala Ala Leu Ile Arg Gly Ile Gln Ser Thr Gly Val Gln Ala Thr
101 130 135 140
102 Ile Lys His Phe Leu Cys Asn Asp Gln Glu Asp Arg Arg Met Met Val
103 145 150 155 160
104 Gln Ser Ile Val Thr Glu Arg Ala Leu Arg Glu Ile Tyr Ala Leu Pro

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/027,000

DATE: 05/13/2002

TIME: 09:51:58

Input Set : A:\GC696-SEQLIST.txt

Output Set: N:\CRF3\05102002\J027000.raw

105				165				170				175				
106	Phe	Gln	Ile	Ala	Val	Arg	Asp	Ser	Gln	Pro	Gly	Ala	Phe	Met	Thr	Ala
107				180					185					190		
108	Tyr	Asn	Gly	Ile	Asn	Gly	Val	Ser	Cys	Ser	Glu	Asn	Pro	Lys	Tyr	Leu
109			195					200				205				
110	Asp	Gly	Met	Leu	Arg	Lys	Glu	Trp	Gly	Trp	Asp	Gly	Leu	Ile	Met	Ser
111		210					215					220				
112	Asp	Trp	Tyr	Gly	Thr	Tyr	Ser	Thr	Thr	Glu	Ala	Val	Val	Ala	Gly	Leu
113	225					230					235				240	
114	Asp	Leu	Glu	Met	Pro	Gly	Pro	Pro	Arg	Phe	Arg	Gly	Glu	Thr	Leu	Lys
115				245						250					255	
116	Phe	Asn	Val	Ser	Asn	Gly	Lys	Pro	Phe	Ile	His	Val	Ile	Asp	Gln	Arg
117			260						265					270		
118	Ala	Arg	Glu	Val	Leu	Gln	Phe	Val	Lys	Lys	Cys	Ala	Ala	Ser	Gly	Val
119			275					280					285			
120	Thr	Glu	Asn	Gly	Pro	Glu	Thr	Thr	Val	Asn	Asn	Thr	Pro	Glu	Thr	Ala
121		290					295					300				
122	Ala	Leu	Leu	Arg	Lys	Val	Gly	Asn	Glu	Gly	Ile	Val	Leu	Leu	Lys	Asn
123	305				310						315				320	
124	Glu	Asn	Asn	Val	Leu	Pro	Leu	Ser	Lys	Lys	Lys	Lys	Thr	Leu	Ile	Val
125				325						330					335	
126	Gly	Pro	Asn	Ala	Lys	Gln	Ala	Thr	Tyr	His	Gly	Gly	Gly	Ser	Ala	Ala
127			340						345					350		
128	Leu	Arg	Ala	Tyr	Tyr	Ala	Val	Thr	Pro	Phe	Asp	Gly	Leu	Ser	Lys	Gln
129			355					360					365			
130	Leu	Glu	Thr	Pro	Pro	Ser	Tyr	Thr	Val	Gly	Ala	Tyr	Thr	Thr	Val	Pro
131		370					375					380				
132	Pro	Ile	Leu	Gly	Glu	Gln	Cys	Leu	Thr	Pro	Asp	Gly	Ala	Pro	Gly	Met
133	385					390					395				400	
134	Arg	Trp	Arg	Val	Phe	Asn	Glu	Pro	Pro	Gly	Thr	Pro	Asn	Arg	Gln	His
135				405						410					415	
136	Ile	Asp	Glu	Leu	Phe	Phe	Thr	Lys	Thr	Asp	Met	His	Leu	Val	Asp	Tyr
137			420						425					430		
138	Tyr	His	Pro	Lys	Ala	Ala	Asp	Thr	Trp	Tyr	Ala	Asp	Met	Glu	Gly	Thr
139		435						440					445			
140	Tyr	Thr	Ala	Asp	Glu	Asp	Cys	Thr	Tyr	Glu	Leu	Gly	Leu	Val	Val	Cys
141		450				455						460				
142	Gly	Thr	Ala	Lys	Ala	Tyr	Val	Asp	Asp	Gln	Leu	Val	Val	Asp	Asn	Ala
143	465				470						475				480	
144	Thr	Lys	Gln	Val	Pro	Gly	Asp	Ala	Phe	Phe	Gly	Ser	Ala	Thr	Arg	Glu
145				485						490					495	
146	Glu	Thr	Gly	Arg	Ile	Asn	Leu	Val	Lys	Gly	Asn	Thr	Tyr	Lys	Phe	Lys
147			500						505					510		
148	Ile	Glu	Phe	Gly	Ser	Ala	Pro	Thr	Tyr	Thr	Leu	Lys	Gly	Asp	Thr	Ile
149			515					520					525			
150	Val	Pro	Gly	His	Gly	Ser	Leu	Arg	Val	Gly	Gly	Cys	Lys	Val	Ile	Asp
151		530				535						540				
152	Asp	Gln	Ala	Glu	Ile	Glu	Lys	Ser	Val	Ala	Leu	Ala	Lys	Glu	His	Asp
153	545					550					555				560	

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/027,000

DATE: 05/13/2002

TIME: 09:51:58

Input Set : A:\GC696-SEQLIST.txt

Output Set: N:\CRF3\05102002\J027000.raw

```

154 Gln Val Ile Ile Cys Ala Gly Leu Asn Ala Asp Trp Glu Thr Glu Gly
155          565          570          575
156 Ala Asp Arg Ala Ser Met Lys Leu Pro Gly Val Leu Asp Gln Leu Ile
157          580          585          590
158 Ala Asp Val Ala Ala Ala Asn Pro Asn Thr Val Val Val Met Gln Thr
159          595          600          605
160 Gly Thr Pro Glu Glu Met Pro Trp Leu Asp Ala Thr Pro Ala Val Ile
161          610          615          620
162 Gln Ala Trp Tyr Gly Gly Asn Glu Thr Gly Asn Ser Ile Ala Asp Val
163 625          630          635          640
164 Val Phe Gly Asp Tyr Asn Pro Ser Gly Lys Leu Ser Leu Ser Phe Pro
165          645          650          655
166 Lys Arg Leu Gln Asp Asn Pro Ala Phe Leu Asn Phe Arg Thr Glu Ala
167          660          665          670
168 Gly Arg Thr Leu Tyr Gly Glu Asp Val Tyr Val Gly Tyr Arg Tyr Tyr
169          675          680          685
170 Glu Phe Ala Asp Lys Asp Val Asn Phe Pro Phe Gly His Gly Leu Ser
171          690          695          700
172 Tyr Thr Thr Phe Ala Phe Ser Asn Leu Ser Val Ser His Lys Asp Gly
173 705          710          715          720
174 Lys Leu Ser Val Ser Leu Ser Val Lys Asn Thr Gly Ser Val Pro Gly
175          725          730          735
176 Ala Gln Val Ala Gln Leu Tyr Val Lys Pro Leu Gln Ala Ala Lys Ile
177          740          745          750
178 Asn Arg Pro Val Lys Glu Leu Lys Gly Phe Ala Lys Val Glu Leu Gln
179          755          760          765
180 Pro Gly Glu Thr Lys Ala Val Thr Ile Glu Glu Gln Glu Lys Tyr Val
181          770          775          780
182 Ala Ala Tyr Phe Asp Glu Arg Asp Gln Trp Cys Val Glu Lys Gly
183 785          790          795          800
184 Asp Tyr Glu Val Ile Val Ser Asp Ser Ser Ala Ala Lys Asp Gly Val
185          805          810          815
186 Ala Leu Arg Gly Lys Phe Thr Val Gly Glu Thr Tyr Trp Trp Ser Gly
187          820          825          830
188 Val

```

191 <210> SEQ ID NO: 3

192 <211> LENGTH: 2502

193 <212> TYPE: DNA

194 <213> ORGANISM: Trichoderma reesei

196 <400> SEQUENCE: 3

```

197 atggctgata ttgatgttga ggccatcttg aagaagctca ccctggccga gaaggctgat      60
198 ctgctggctg gtatogactt ctggcacaca aaggctctcc ccaagcatgg agtccccctct      120
199 ctccgcttta cagatggccc caacggcgta agagggacca agttcttcaa tggcgctccct      180
200 gcggcctgct tcccttgccg cactgcgtc ggttccacat tcaaccaaac tctgctcgaa      240
201 gaggcaggta agatgatggg caaagaggcc atcgctaaga gtgcgcatgt gatcctcggc      300
202 ccgactatca acatgcaacg ctcccctctc ggtggacgtg gcttcgagtc gattggtgag      360
203 gatccgttcc tggcgggcct gggagctgcg gctctcatcc gcggcattca gagcactgga      420
204 gtgcaggcta cgatcaagca cttttgtgc aatgatcagg aggacaggcg catgatggtg      480
205 cagagcatcg tcacggagcg ggctctccgt gaaatctacg cactcccggt ccagattgct      540

```

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/027,000

TIME: 09:51:59

Input Set : A:\GC696-SEQLIST.txt

Output Set: N:\CRF3\05102002\J027000.raw

206	gtgcgagact	cccagccggg	tgcgttcattg	acggcgtaca	atggcatcaa	tggcgtgtcg	600
207	tgcagcgaga	accctaaata	tcttgatggg	atgcttcgaa	aggaatgggg	ttgggatggc	660
208	ctaatacatga	gcgactggta	cggcacatac	agtaccacag	aagccgttgt	ggcaggcctc	720
209	gacctcgaga	tgcccggacc	tccacgcttc	cgaggagaaa	cactcaagtt	caacgtctcc	780
210	aacggaaagc	cctttatcca	cgtcattgac	cagagggcta	gggaagttct	tcagttcgtc	840
211	aagaagtgtg	ctgcctccgg	agtgcaggag	aacggccccc	agacgactgt	caacaacacc	900
212	cccgaacagg	cagctctcct	ccggaaggtt	ggcaacgagg	gcacgtgtgt	gctgaagaac	960
213	gagaacaacg	ttctgccctt	gagcaagaag	aagaagacgc	tgattgtcgg	ccccaacgcc	1020
214	aagcaggcca	cataccacgg	cggaggctct	gccgcactca	gggcctacta	cgagtcact	1080
215	ccctttgacg	gcctcagcaa	gcagctcgag	acgcccgcct	cgtacaccgt	cggcgcctac	1140
216	accaccgttc	ctcccattct	aggcgagcag	tgcctcacgc	ccgacggcgc	tccgggcatg	1200
217	cgctggaggg	tcttcaacga	gccccctgg	acccctaacc	gccagcacat	tgacgagctc	1260
218	ttcttcacca	agacggacat	gcacctggtg	gactactacc	accccaaggc	ggcagacacg	1320
219	tggtacgccg	acatggaggg	cacgtacacc	gccgacgagg	actgcacctc	cgagctcggc	1380
220	ctcgctgtct	gcggcacggc	aaaggcgtac	gtagacgacc	agctcgtcgt	cgacaacgcc	1440
221	accaagcagg	tccccggcga	tgccttcttc	ggctccgcc	cccgcgagga	gacgggcccgc	1500
222	atcaatctcg	tcaagggcaa	cacgtacaag	ttcaagatcg	agttcggtc	cgcaccacc	1560
223	tacaccctca	agggcgacac	catcgtcccc	ggccacggct	ccctccgct	cggcggctgc	1620
224	aagggtcattg	acgaccaggc	cgaaatcgaa	aagtcgctcg	ccctcgccaa	ggagcacgac	1680
225	cagggtcatca	tctgcgcggg	ccttaacgcc	gactgggaga	ccgagggcgc	cgaccgcgcg	1740
226	agcatgaagc	tccccggcgt	gctggaccag	ctcattgccg	acgtggccgc	cgcgaaccca	1800
227	aacaccgtcg	tcgtcatgca	gacgggcacc	cccaggagga	tgccctggct	cgacgccacg	1860
228	cccgcctca	tccaggcctg	gtacggcggc	aacgagacgg	gcaactccat	tgccgacgtc	1920
229	gtctttggcg	actacaaccc	ctcgggcaag	ctgtccctca	gcttccccaa	gcgcctgcag	1980
230	gacaaccccg	cgtttctcaa	cttccgcacc	gaggccgggc	gcacgctgta	cggcgaggac	2040
231	gtctacgtcg	ggtacaggta	ctacgagttt	gccgacaagg	acgtcaattt	cccctttggc	2100
232	cacggcctgt	cctacaccac	ttttgccttt	tccaatctct	ccgtgtctca	caaggacggc	2160
233	aagctgagcg	tgccccctct	cgtgaagaac	acgggtcccg	tgcccggcgc	acaggtggcc	2220
234	cagctctacg	tcaagccctt	ccaagcggcc	aagattaacc	gccccgtcaa	ggagctcaag	2280
235	ggcttcgcaa	aggtcgaaact	gcagcccggc	gagacgaagg	cggtgacaat	cgaggagcag	2340
236	gagaagtaag	tcgctgcgta	ttttgatgag	gagcgggac	agtgggtgtg	cgaaaaggg	2400
237	gactatgagg	ttatcgtgag	cgacagcagc	gcagcgaagg	atgggggtgc	gctcaggggt	2460
238	aagtttacgg	tgggagagac	gtattgggtg	tctggcgtgt	aa		2502

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/027,000

DATE: 05/13/2002

TIME: 09:52:00

Input Set : A:\GC696-SEQLIST.txt

Output Set: N:\CRF3\05102002\J027000.raw